From Raw Data to Physics Results Q&A

Paul Laycock

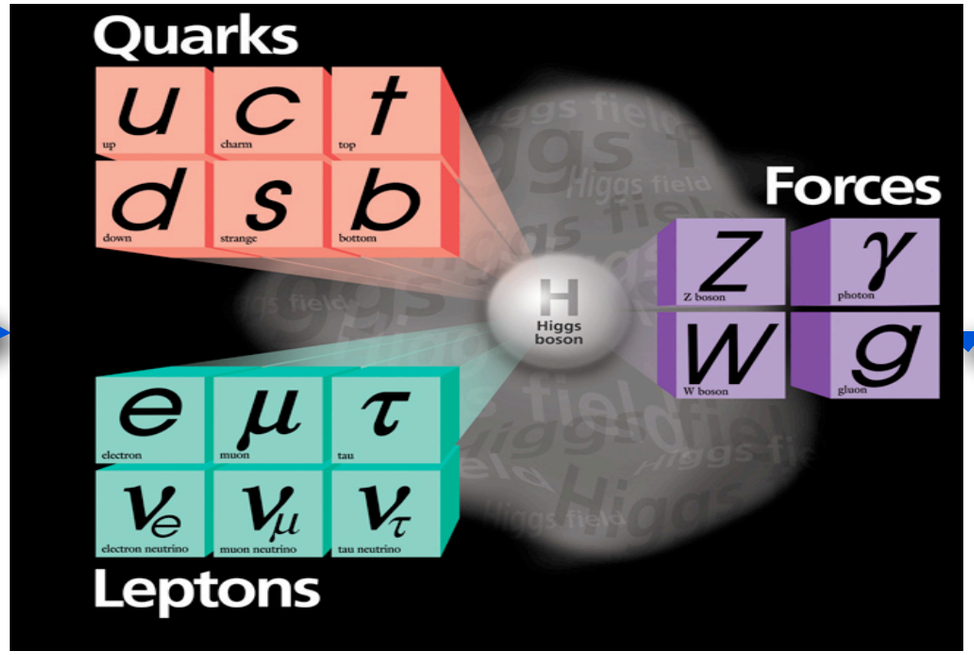
BROOKHAVEN NATIONAL LABORATORY

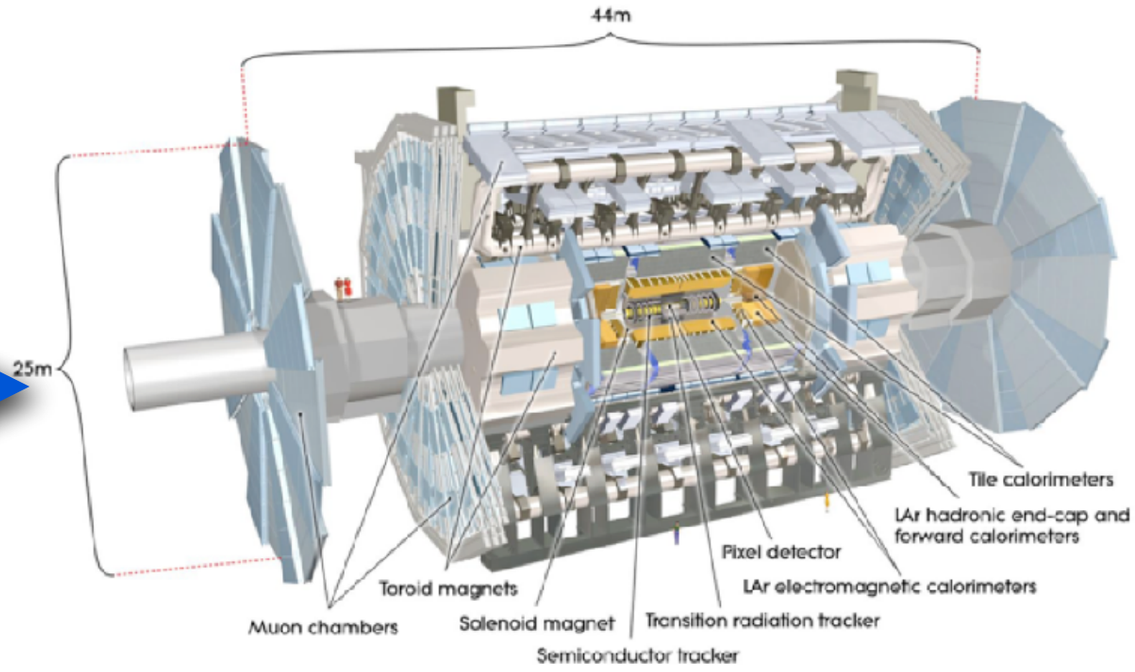
U.S. DEPARTMENT OF ENERGY

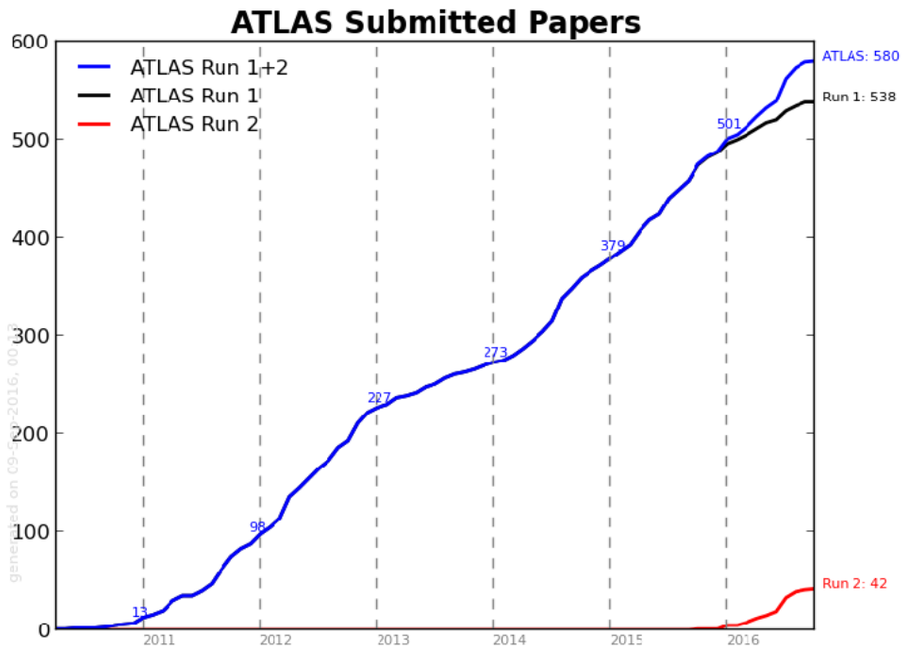
Questions - 1

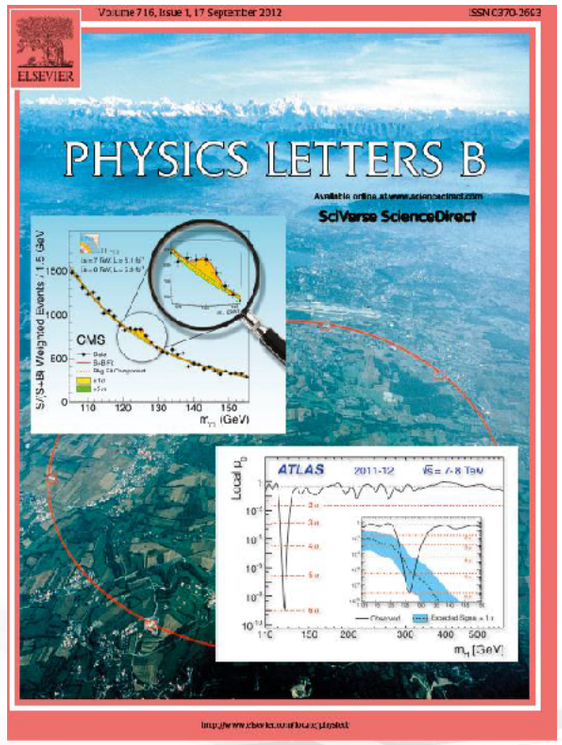
What happens to the large amount of data after analysis? Is it just deleted or is it stored anywhere for future reference?

The physics cycle

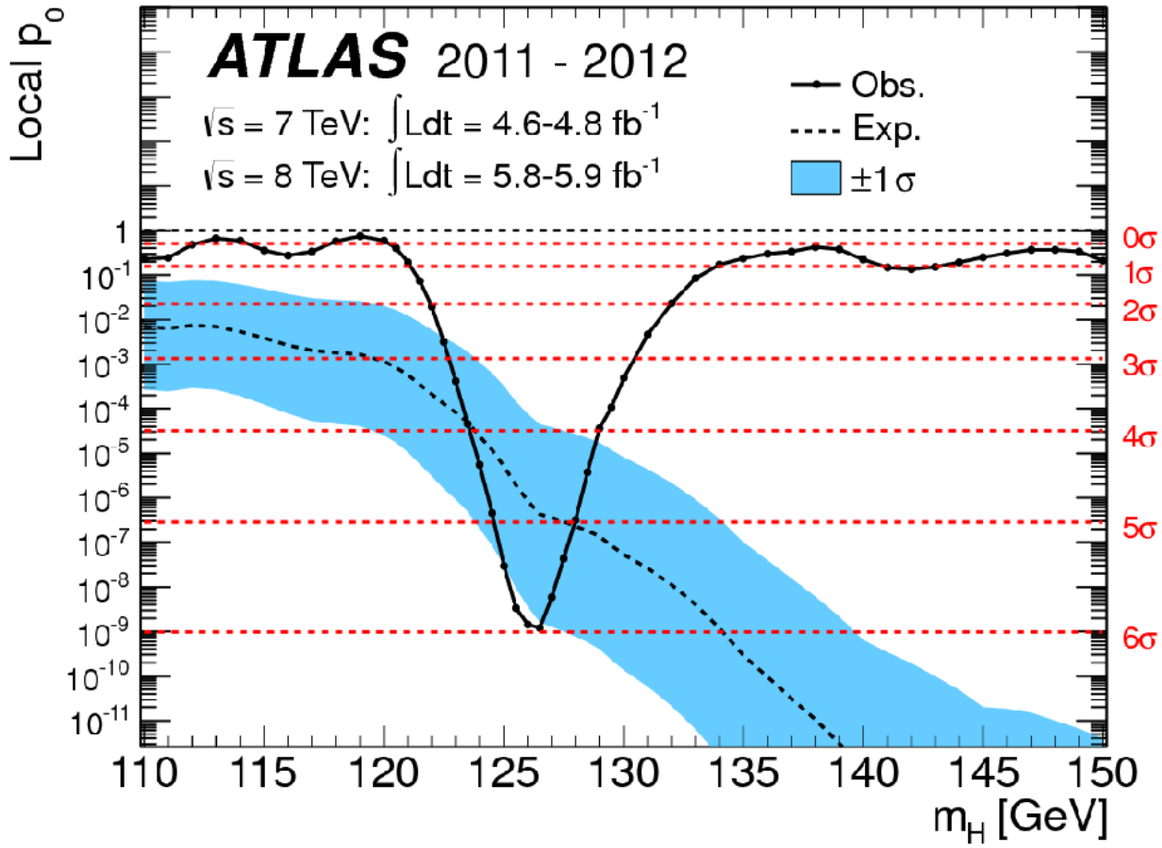








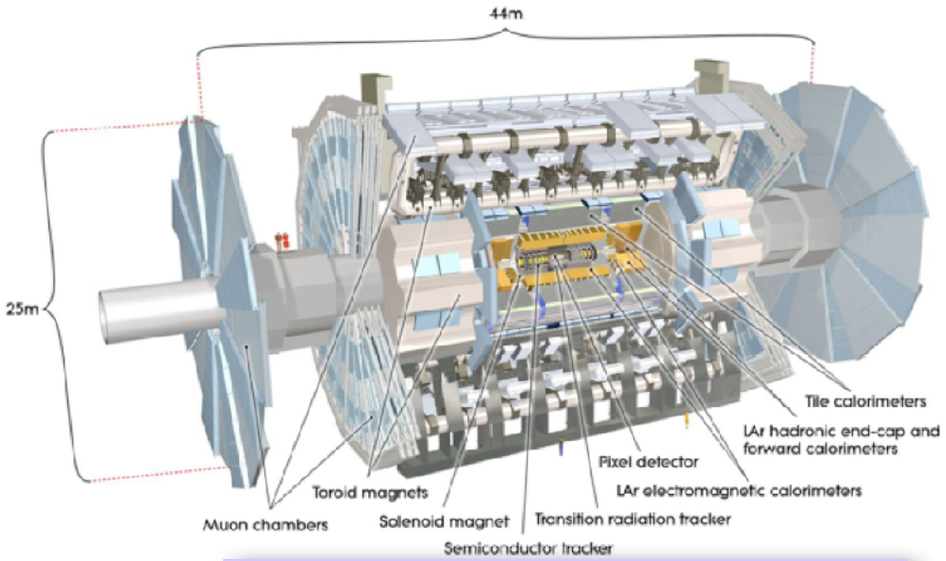
Higgs discovery in 2012



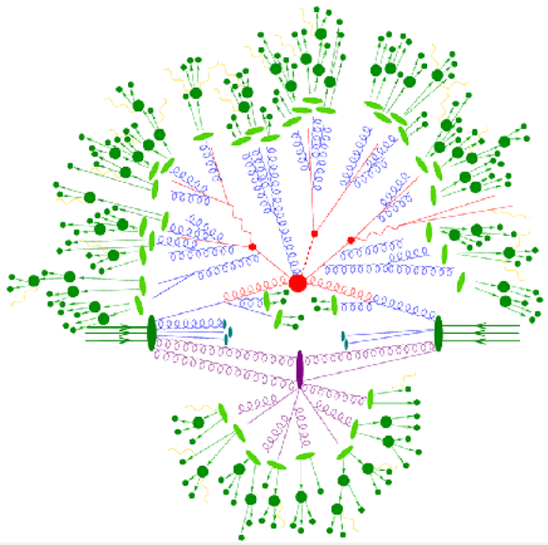
- In 2012 the number of observed events (**6σ**) was consistent with, and in excess of the number of events expected for a standard model Higgs (**5σ**)

- ***Question*** *- Imagine we had several more Large Hadron Colliders, with a total of 9 independent measurements possible. Roughly how many measurements would you expect to lie* ***outside*** *the ±1σ blue band?*

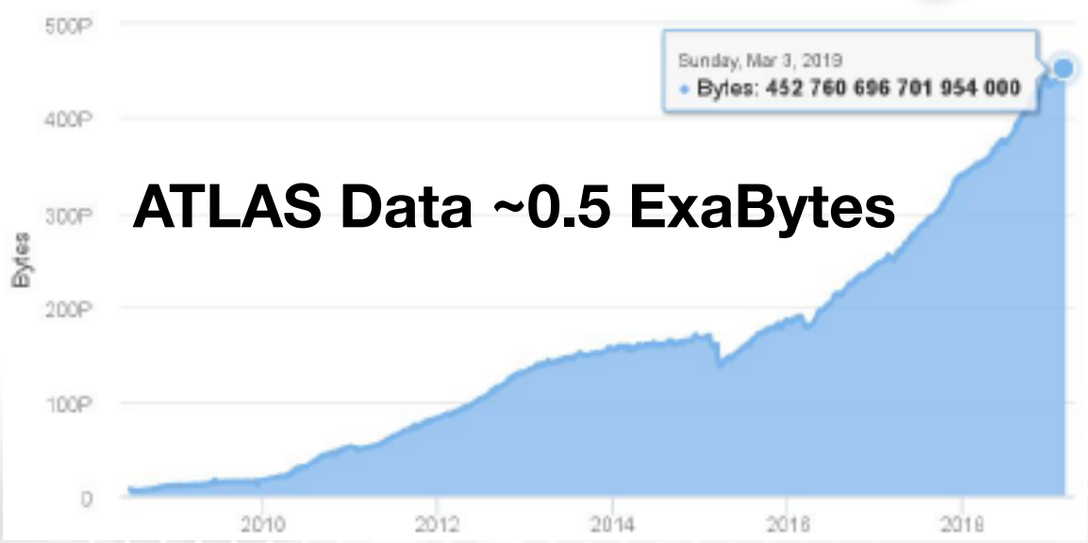
Exabyte-scale physics analysis



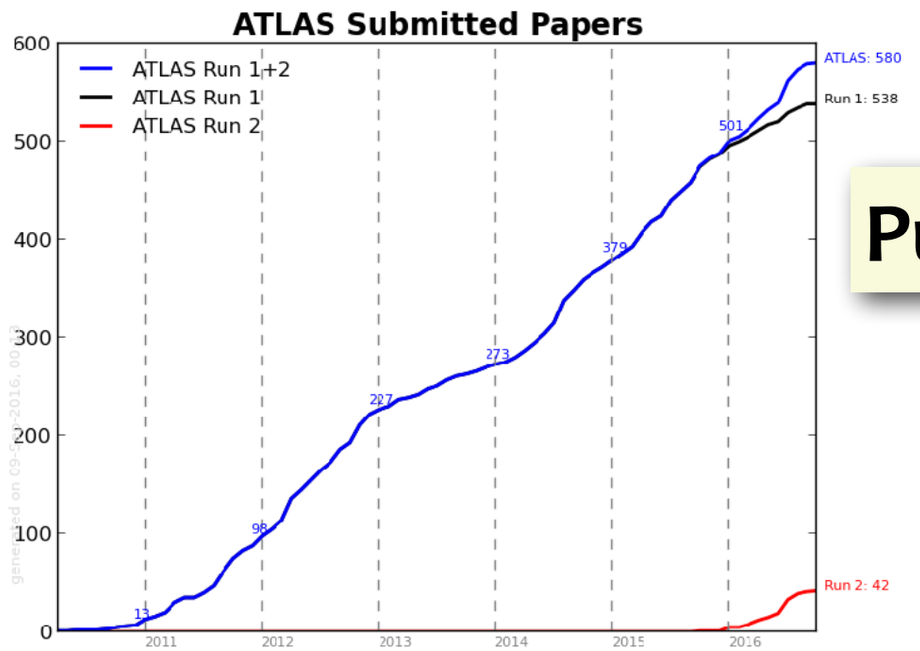
Exabytes of Data



Exabytes of Simulation

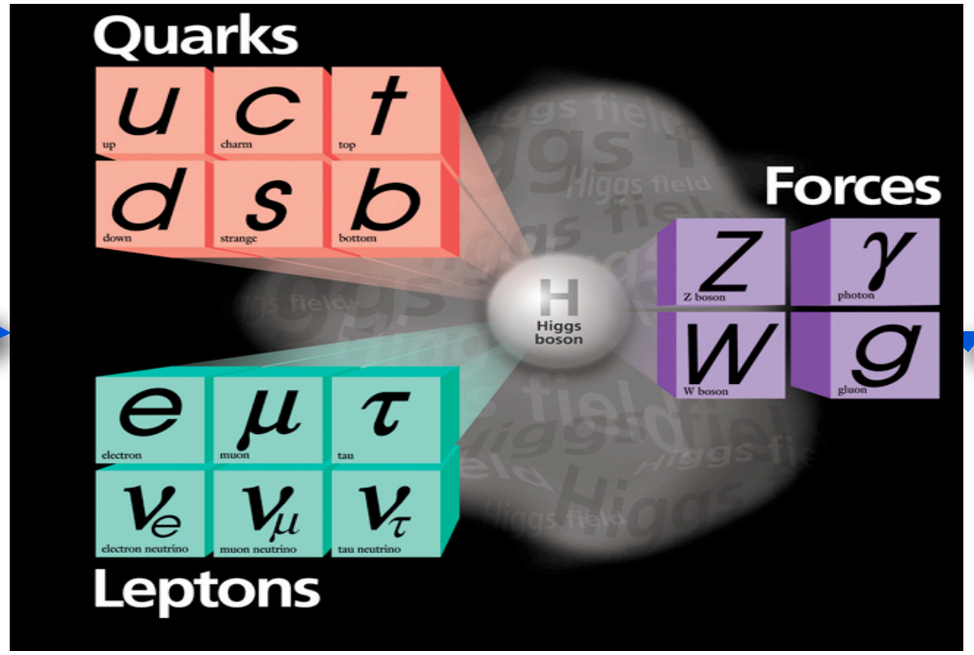


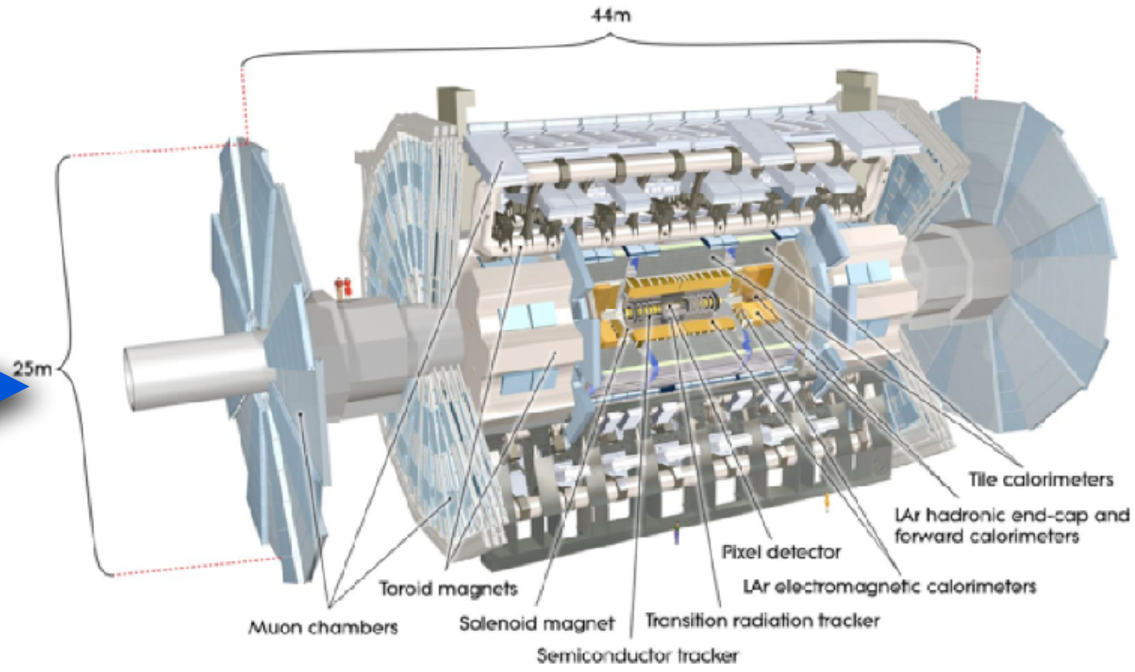
ATLAS Data ~0.5 ExaBytes

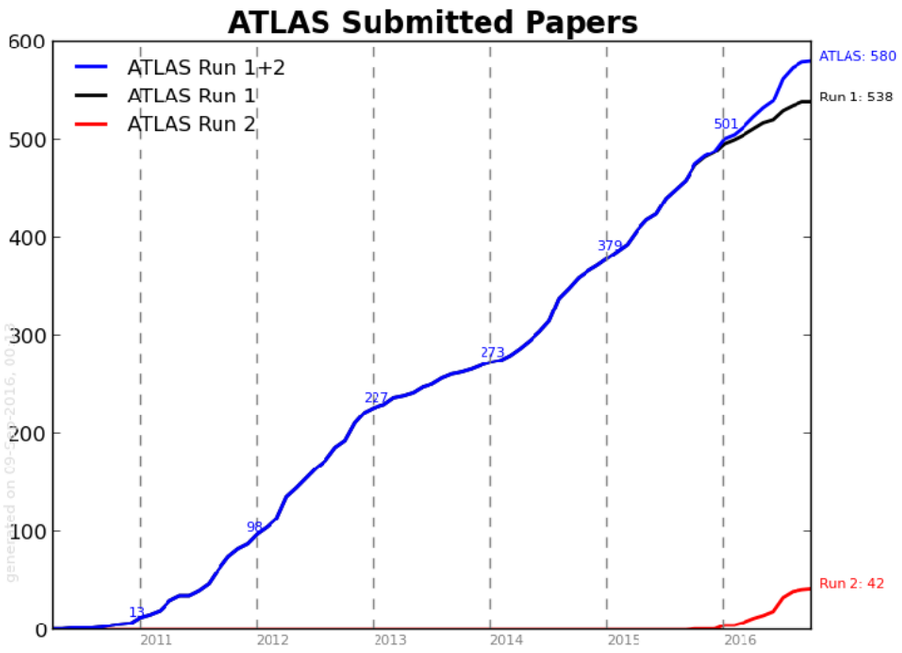


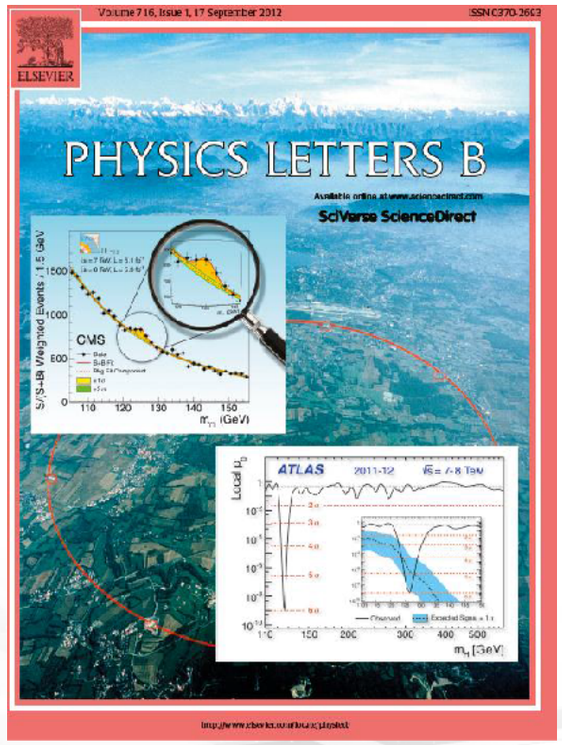
Publish!

The physics cycle









Questions - 1

What happens to the large amount of data after analysis? Is it just deleted or is it stored anywhere for future reference?

but what happens after the experiment ends??

Data Abstraction Hierarchy

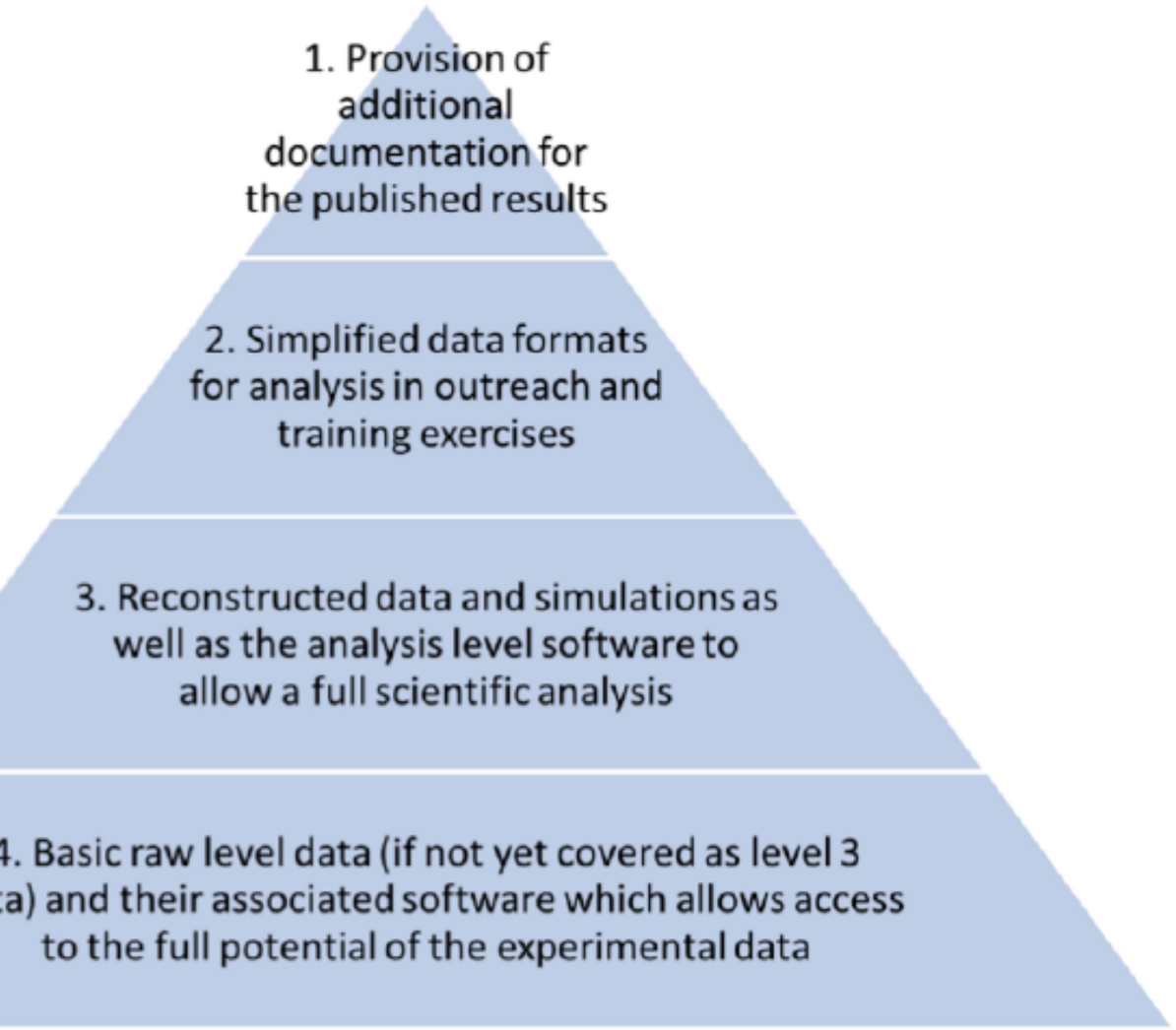
~KB/paper

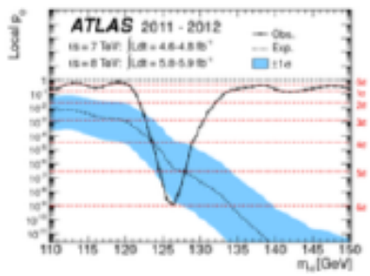
~GB/analysis

~TB/analysis

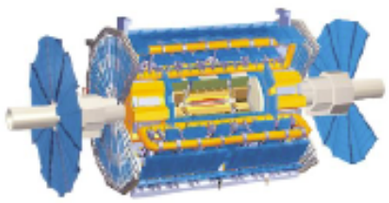
~PB/year

~GB/sec





analysis



Four data levels for capture, preservation and opening

adapted from slide by Tibor Simko

Making data public

CERN Open Data

Status: *production* (since November 2014)

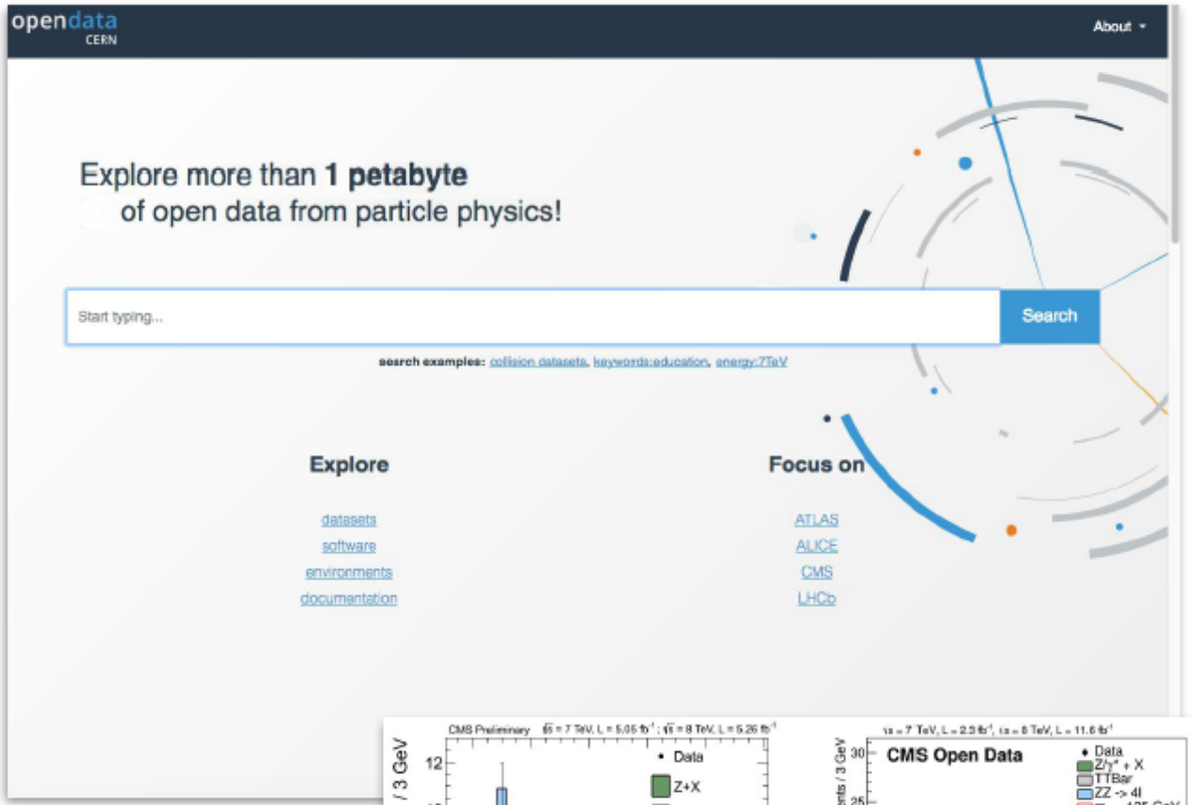
Size: 7K records, 800K files, 2 PB size

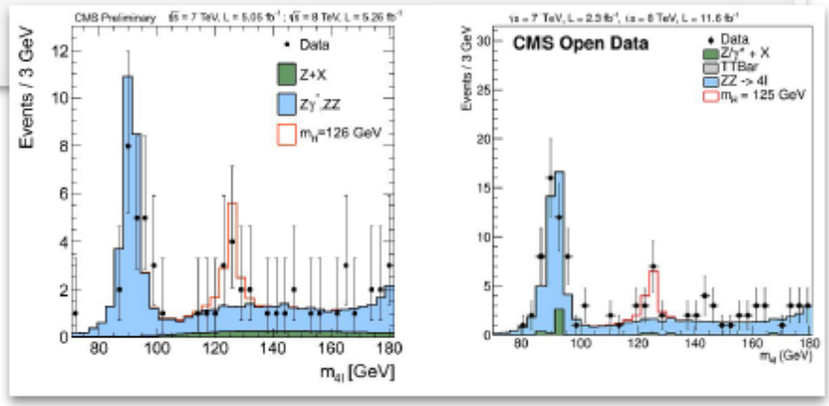
Purpose: "big data" sharing of event-level particle physics data and accompanying code for both education and research purposes

Content: raw samples, collision & simulated & derived datasets, docs, configs, software tools, example analyses, VMs, event display

Community: ALICE, ATLAS, CMS, LHCb, OPERA (coming: JADE, Data Science)

Notes: independent expert curation; batch ingestion workflows with Collaborations





Slide: Tibor Simko http://opendata.cern.ch

Questions - 2

What will happen in the future if we have more experiments that need high quantities of data and not enough computing power?

Experimental Physics at the LHC

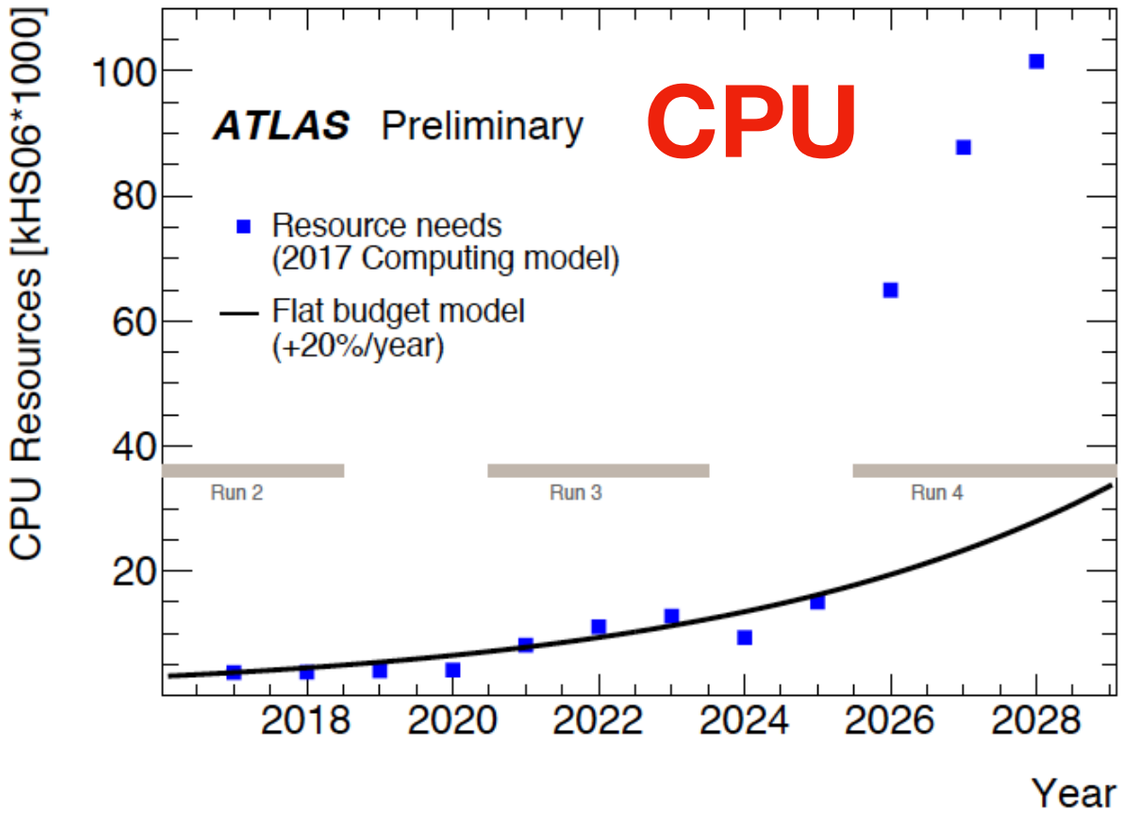
- To measure rare processes you need high statistics
- Measuring a 1 in a billion process means you'd better have at least a billion events
- Measuring with high precision requires even higher statistics

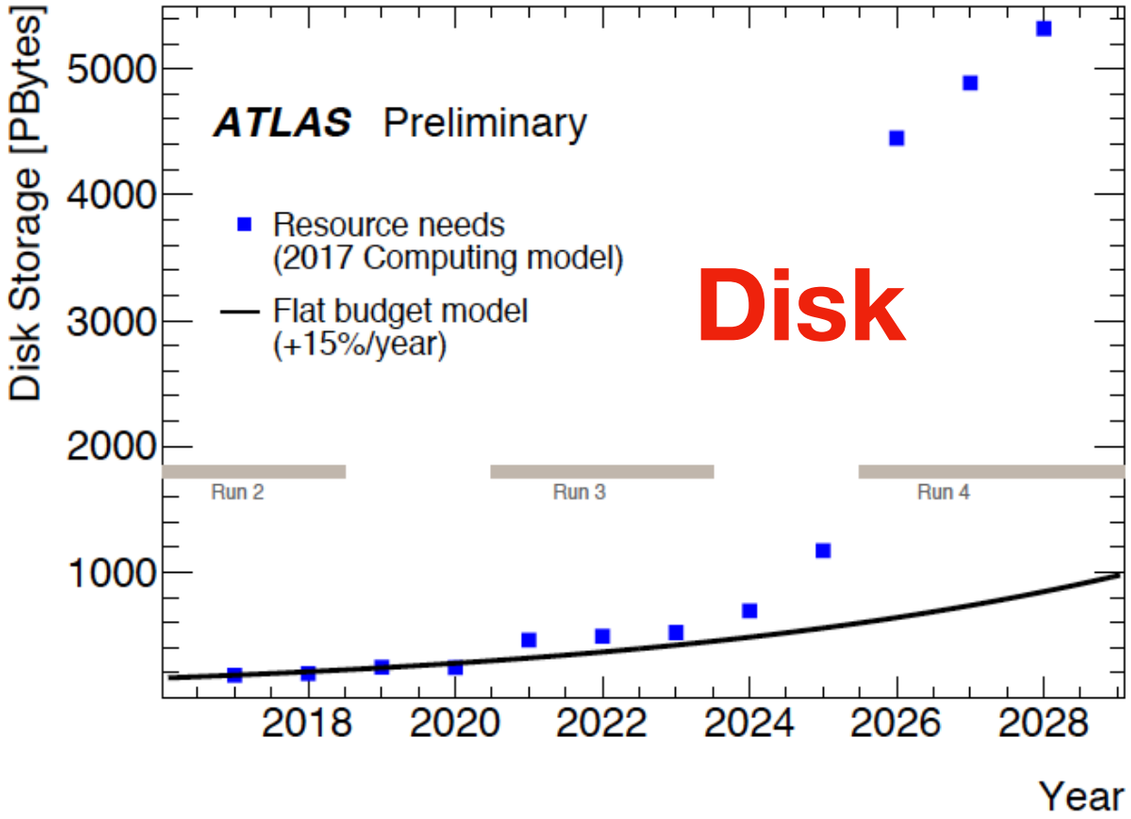
The solution - increase the intensity of particle beams, this increases the interaction rate. In parallel also increase the rate of data that record.

- Distinguishing signal from background requires high granularity detectors: lots of information for each event

Recorded rate * detector event size = Data volume
Data rate * data size = Data volume

LHC future computing





The computational needs of the high luminosity LHC will far outstrip our ability to process it

We need solutions to these problems!